Applying DC to Institutional Data Repositories

Robin Rice
EDINA and Data Library
University of Edinburgh, UK
R.Rice@ed.ac.uk

Keywords: data curation; datasets; DDI; DSpace; institutional repositories; Dublin Core

DISC-UK DataShare (2007-2009)[73], a project led by the University of Edinburgh and funded by JISC (Joint Information Systems Committee, UK), arises from an existing consortium of academic data support professionals working in the domain of social science datasets (Data Information Specialists Committee-UK). We are working together across four universities with colleagues engaged in managing open access repositories for e-prints. Our project supports 'early adopter' academics who wish to openly share datasets and presents a model for depositing 'orphaned datasets' that are not being deposited in subject-domain data archives/centres.

Outputs from the project are intended to help to demystify data as complex objects in repositories, and assist other institutional repository managers in overcoming barriers to incorporating research data. By building on lessons learned from recent JISC-funded data repository projects such as SToRe[74] and GRADE[75] the project will help realize the vision of the Digital Repositories Roadmap, e.g. the milestone under Data, "Institutions need to invest in research data repositories" (Heery and Powell, 2006).

Application of appropriate metadata is an important area of development for the project. Datasets are not different from other digital materials in that they need to be described, not just for discovery but also for preservation and re-use. The GRADE project found that for geo-spatial datasets, Dublin Core metadata (with geo-spatial enhancements such as a bounding box for the 'coverage' property) was sufficient for discovery within a DSpace repository, though more in-depth metadata or documentation was required for re-use after downloading. The project partners are examining other metadata schemas such as the Data Documentation Initiative (DDI) versions 2 and 3, used primarily by social science data archives (Martinez, 2008). Crosswalks from the DDI to qualified Dublin Core are important for describing research datasets at the study level (as opposed to the variable level which is largely out of scope for this project).

DataShare is benefiting from work of of the DRIADE project (application profile development for evolutionary biology) (Carrier, et al, 2007), eBank UK[76] (developed an application profile for crystallography data) and GAP[77] (Geospatial Application Profile, in progress) in defining interoperable Dublin Core qualified metadata elements and their application to datasets for each partner repository. The solution devised at Edinburgh for DSpace will be covered in the poster.

References

Carrier, Sarah, Jed Dube, and Jane Greenberg. (2007). The DRIADE project: Phased application profile development in support of open science. *Proceedings of the International Conference on Dublin Core and Metadata Applications, 2007*, (pp. 35-42).

Heery, Rachel, and Powell, Andy. (2006). *Digital repositories roadmap: Looking forward*. Retrieved from http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/#roadmap-200604.

Martinez, Luis. (2008). *The Data Documentation Initiative (DDI) and institutional repositories*. Retrieved from http://www.disc-uk.org/docs/DDI_and_IRs.pdf.

---

[73] http://www.disc-uk.org/datashare.html
[74] http://www.era.lib.ed.ac.uk/handle/1842/1412
[75] http://edina.ac.uk/projects/grade/
[76] http://www.ukoln.ac.uk/projects/ebank-uk/schemas/profile/
[77] http://edina.ac.uk/projects/GAP_summary.html